A schema for cryptographic keys generation using hybrid biometrics

Conference or Workshop Item

A schema for Cryptographic Keys generation using Hybrid Biometrics

Kiran Voderhobli, Dr. Colin Pattinson, Dr. Helen Donelan
Kiran_voderhobli@yahoo.co.uk, c.pattinson@leedsmet.ac.uk, h.donelan@leedsmet.ac.uk
Leeds Metropolitan University,
Leeds, United Kingdom, LS6 3QS

Abstract - **Biometric identifiers refer to unique physical properties or behavioral attributes of individuals. Some of the well known biometric identifiers are voice, finger prints, retina or iris, facial structure etc. In our daily interaction with others directly or indirectly, we implicitly use biometrics to know, distinguish and trust people. Biometric identifiers represent the concept of "who a person is" by gathering vital characteristics that don't correspond to any other person. The human brain to some extent is able to ascertain disparities or variation in certain physical attributes and yet verify the authenticity of a person. But this is difficult to be implemented in electronic systems due to the intense requirements of artificial decision making and hard-coded logic.**

This paper examines the possibility of using a combination of biometric attributes to overcome common problems in having a single biometric scheme for authentication. It also investigates possible schemes and features to deal with variations in Biometric attributes. The material presented is related to ongoing research by the Computer Communications Research group at Leeds Metropolitan University. We use this paper as a starting step and as a plan for advanced research. It offers ideas and proposition for implementing hybrid biometrics in conjunction with cryptography. This is a work in progress and is in a very preliminary stage.

I. INTRODUCTION

Currently, biometrics is used to grant access of some kind, like verified and logged entry into buildings. The complication of implementing accurate recognition schemes has restricted the use of biometrics for financial transactions. In other words, the preciseness of the system is dependent on how well the digital system is able to match a user's physical properties under varying conditions. In this paper we conduct a brief study of how accurate biometric identifiers are, and the suitability in implementing them for a robust authentication system.

Depending on the nature of the application, a single biometric identifier would often be inadequate to create a significant level of security.

For example, this has been stressed in the past with regard to the customer base of financial institutions [1]. The reason for this is that all customers might not be able to produce a given biometric attribute due disabilities or environmental factors.

This major problem could be alleviated by using a collection of biometric attributes, rather than a single one. The degree of uniqueness and robustness against circumvention [2] makes biometrics ideal for applications involving cryptography. We will need to devise a mechanism that uses a set of characteristics but at the same time ensure that it caters and adapts to different conditions.

II. ISSUES IN HYBRID BIOMETRICS

In a real-time scenario, from a set of ***B*** biometric attributes, a person may not be able to produce a subset of attributes. We will need to deal with such situations and clearly define the acceptable level of identification. For example, assume an application defines ***B={ Iris, fingerprints, voice, password, face }***. If on any one particular instance, an authentic person may not be able to produce all the attributes of ***B*** but rather a subset ***S={ Fingerprints, face, password }***.This could be due to change in physical attributes of a person, or due to external influential factors. Correlating, ***S*** with ***B*** is a major challenge. Incomplete and erroneous input must be distinguished. A genuine person might furnish incomplete biometric data. In such cases the system must decide if the identification process has sufficient information to authenticate a person. This is extremely critical in those cases when hybrid biometric data is used for key generation that is used with standard crypto-algorithms.

We examine the following issues

- Extracting and representing biometric data that could for instance be used for hybrid biometric key generation.
- Dealing with erroneous and incomplete verification data

ISBN: 1-9025-6013-9

Although usage of biometric data in cryptosystems is new, successful attempts in implementation have been made in the recent past. One such system is the BIO-IBS, which is short for biometric Identity Based Signature scheme [3]. The system, effectively deals with changes to biometric data, which vary over time. Davida et. al. [4][5] created a system which uses a biometric attribute as a key. Another major research was by F. Hao and C.W. Chan [6] who presented a method to generate keys from handwritten signatures.

III. EXTRACTING BIOMETRIC DATA

The typical setting of a hybrid biometric authentication system will be wide spread and include individual sub-systems to derive vital data from some specific user attribute. The sub-systems will need a means of representing physical features and further also a method to compare with the template. The likely presence of noisy data is an extra complexity at each sub-system. The first step involved will be to compare variations to detect noisy data.

In the case of a hybrid mechanism, template matching for each sub-system is only a part of the overall process. At a higher level the representation and template comparison must be used to consolidate distributed results. The major template will be the key that is defined for the user and has to match the aggregated key formation derived during authentication. An intermediate comparator key can be generated based on some function imposed on independent approximations obtained from various sub-systems. The sub-systems may be physically located in a single module, but they have well defined unit of task associated with a definite attribute. A sub-system could be a device like the finger print scanner, iris scanner etc.

The scheme we present basically has four important phases (Fig. 1)

- Providing enrolment information.
- Compare – Detecting noisy data
- Approximation – Independent representation of attributes derived from noisy data
- Consolidate – Creating rigid comparator key from approximation.

It could be noted that the consolidate function is first applied per individual at the enrolment process. Subsequently, it is applied for every authentication session. Complex consolidate functions could also embed extra security transformations to resist attacks. For example it could add extra bits to prevent key generation from being straight-forward.

Let us first see how biometric data can be represented. In a recent publication Dodis et. Al. [7], described different mechanisms for representation. Even the BIO-IBS mentioned before [3], uses a significant part of that research.

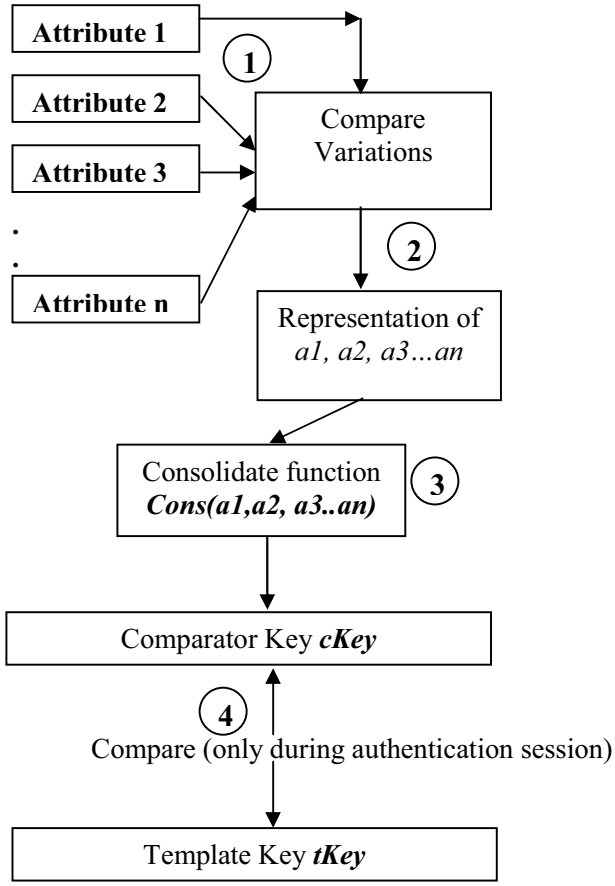

Fig. 1 Enrolment and hybrid key generation

IV. REPRESENTING ATTRIBUTES

The material presented below is based on the research of Dodis et Al [7], and we observe how the methodology could be used in conjunction with hybrid biometrics. We only briefly explain those concepts that are relevant to our research.

As said in the previous section, the representation scheme must address the possibility of noisy data. A combination of secure-sketch and fuzzy extractors can be used to encompass handling representation, biometric matching and to deal with noisy data.

Every biometric attribute is assumed to have a metric space which is normally infinite. However, to restrict the study to a well defined set we could assume the metric space to be finite. This will help us to find out the magnitude of difference between any two elements of the metric space. Each element is a

representation of some Biometric input. Let ***M*** be the metric space for any one type of Biometric input.

A secure sketch ***SS*** can be defined on ***M***. A secure sketch is a randomized map. Let us define a secure sketch ***SS (w)***, where ***w*** is some (possibly constant) representation. Using ***SS(w)*** and some deterministic recovery function **Rec** it is possible to derive ***w*** from some ***w'*** close to ***w***. This also takes into account the distance between ***w*** and ***w'***. In others words it is possible to define an acceptable variation level between two values belonging to ***M***. The derivation of ***w*** will fail if ***w-w'>vl***, where ***vl*** is the variation allowed. ***w-w'*** is the distance measured either based on hamming metric, set difference metric or edit metric. Ref. [7] will provide a detailed understanding of these metrics. In simple terms, these metrics are used to evaluate the variation in the number of bit positions of two given inputs.

A secure sketch can be extended to create fuzzy extractor. A fuzzy extractor on a metric space ***M*** is defined by two procedures **Gen** (probabilistic generator), **Rep** (reproduction procedure). The generator function accepts an input ***w*** and outputs an extracted string ***R*** and also a public string ***P***. The reproduction procedure uses ***P*** and any input ***w'*** close to ***w***, to extract ***R***. The input ***w*** mentioned here is the biometric attribute presented during the enrollment process. The generation of ***R*** and ***P*** is dependent on the logic built into the **Gen** function.

Since the above scheme yields a definitive ***R*** for a given input, this method could be exploited to generate identifiers for creating stored biometric templates. The identifier (or a part of it) could then be taken as it is or further processed to form a key. Let us perform a run through the enrollment process (*step 1 in fig. 1*). This is the phase where a user registers, by presenting biometric data. Our actual scenario includes a multitude of biometric attributes, but for the sake of simplicity at this stage, we will isolate the case for a single attribute. Let us consider voice. When a new user is registering, he will read out a password into a microphone. This will be some ***w*** belonging to the valid metric space ***M*** and that can be represented using ***n*** bits. The fuzzy extractor invokes the **Gen** function with the input ***w***. **Gen(*w*)** will produce some ***R*** and a public string ***P***. The ***R*** that is obtained here can be used as a partial template or can be altered based on some predefined function. As said before, this can be used as a (or a part of) main key (see *section 5* on consolidation). We have assumed that only one voice sample is collected. Alternatively, a series of samples maybe collected and a mean of these samples based on variations between each other could be used to determine ***w***.

For the attributes representation stage (*step 2 in fig. 1*), the ***R*** values of each attribute is considered as the building blocks of the main key. For example attribute ***a1*** is represented by ***R1***.

The consolidate function is executed both during enrolment and at every authentication session. The consolidate function could be some hashing function. The (one-time) enrolment session uses this function to generate the full key. The authentication sessions use the function to generate a comparator key. An authentication session grants a successful entry if the comparator key matches the already stored full key. For this to happen, every biometric attribute with possible input ranges ***{w'1, w'2, w'3… , w'n}***, all sufficiently close to ***w***, must use the recovery function (**Rec**) and the corresponding public string to regenerate **R**. When the recovery functions yield a successful result for each and every biometric attribute, we are sure to have a key match. This is because the input to the consolidate function during the enrolment and authentication has matched and the derived comparator key and the stored key are the same.

While the consolidate function can impose constraints for making a key complex, the tolerance level for each biometric attribute can be defined at the top most (end-user) level. The tolerance levels can be independently set based on the set of small variations on the metric space. Let us again recollect the condition ***w-w'<vl***. Defining the range of ***vl*** determines the tolerance level. If ***vl*** is from a very small range, the system is very precise and more tolerant. A significantly large range of ***vl*** will make the system more flexible by allowing a larger degree of erroneous data.

V. CONSOLIDATION

The consolidation function will involve some function to further process the attributes. The function would take as parameters, biometric representations for ***n*** physical attributes to yield a key. During the enrolment process, the consolidation yields a template that is permanently stored. The consolidation is also executed for every authentication session, and each time it outputs a comparator key (***cKey***). This key could be used for encrypting real-time communications, if and only if it exactly matches the stored template key (***cKey=tkey***).

The consolidation function can be custom formulated in any appropriate way to offer maximum security by adding complexity to the key bit sequence. But in our research, we have considered using a concept that is quite similar to hashing [8]. A typical hash function will map large domains to smaller ranges using many to one function. We know this will have the effect on the size of the overall key. So large biometric representations can be collected and collated together to be mapped into a fixed length key domain. The main criteria we use for our consolidate function is that it should be a one way function that performs some sequence alterations and/or bit additions to the initial key bit sequence. The idea is to generate a key whose bit sequence is so large that it can resist any brute force attack. If the initial representation of each of the attributes is many bits long, the

overall summation of various attributes will be resistant to brute-force attacks even without bit padding. Padding bits is useful, when fewer attributes are provided to authenticate and to generate fixed length large keys.

The consolidate routine requires an initial bit sequence that is obtained from the ***R*** representation values of each attribute. Over ***n*** biometric attributes we will have a set of R_i where **I Є {1..n}**. Each of the R_i might be subject to initial bit padding to attain a fixed size sequence. The padding will be in accordance to the expectations of the length of the output key. Now we have a finite length bit string say ***S***. ***S*** will be the input for the one way function that will output the key. The string is divided into blocks. The function will perform a well-defined set of iterations and bit-wise processing on each block of ***S***.

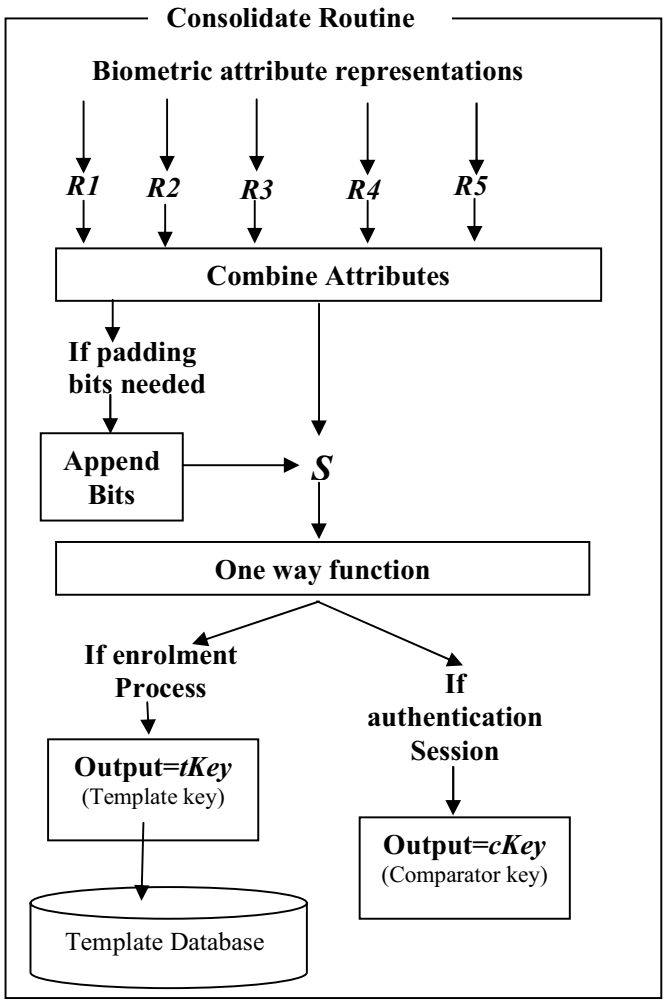

Fig 2.Consolidate routine

At the enrolment session for a user the consolidate routine, creates and stores a key in the template database, which is again used by the same routine during authentication to compare a generated string (***cKey***) to the stored key (template).

VI. DEALING WITH PARTIAL ATTRIBUTE SUBMISSIONS

Biometrics is subject to inconsistent presentation that depends on how the attribute is presented and represented. Often, it would not be possible for a valid user to present all attributes accurately which results in a major difference to biometric data presented during enrolment and some instance of authentication. Take for instance, a user is suffering from a cold and there is a large variation in the voice sample provided at authentication. If w_{voice} is the voice sample provided during enrolment and w'_{voice} is the voice sample provided when the user has a cold, then there could be a possibility of $w_{voice} - w'_{voice} > vl_{voice}$. When this is the case, recovering R_{voice} using secure sketch would not be possible and authentication will fail.

The former method presented creates a template that is stored permanently and subsequently used for comparisons. But would it be possible to generate keys at run-time based on partial attribute submission? The challenge here is to permit access to a person and release a suitable key for transactions, based on a fair amount of certainty determined by fewer biometric attributes. As a part of our research in the lab, we are currently endeavoring to implement such a test-schema for a PKI environment. Obviously, the key(s) here will not be constant and will vary for each session, depending on which attributes and how many attributes are presented. We have adopted a very basic approach that is elementary at the initial stages of the research.

The system hence uses partial consolidation. This means that if ***n*** attributes are presented at an authentication session, and if only some of the ***n*** attributes are known to match, then the system ignores erroneous attributes and consolidates only correct ones. Therefore, an intermediate comparator key cannot be matched to some stored template as it may differ for each session. The template key created in this scheme is not used for comparison but rather released immediately, if partial matches are obtained.

It is assumed that during enrolment a user is able to provide all specified biometric attributes. Let us say the enrolment procedure demands fingerprints, iris, voice, facial feature and password. At enrolment the representations for all these attributes are generated. Hence we have representations R_{voice}, R_{iris}, $R_{fingerprints}$, R_{facial} and ***password***. These representations are utilized to create a key using a one way function.

There are a couple of constraints we need to impose to ensure that the system does not compromise on security given that a

person can gain entry just based on fewer attributes. For instance the administrator can fix that at least three out of a five attributes are mandatory. He can also attach weights to each of the attributes and designate that the summation of the weights of correctly produced attributes is greater than some value. Let us for example use the following weights

Voice= 7, Iris=4, Facial=10, Fingerprints=3, Password=16

Let the designated summation of weights be 15. Also, let number of mandatory attributes is set to 3. Suppose during an authentication session, only 3 of the five attributes, voice, iris and fingerprints are found to be correct. Although this satisfies the number of mandatory attributes, it does not satisfy the constraint that the summation of the weights must be greater than or equal to 15. Again, if the system is successful in matching only voice and password, the authentication will fail, even though the summation of weights is greater than 15. A valid authentication combination would be voice, iris and password.

Each of the attributes submitted during authentication will be checked for $R_n - R'_n < vl_n$ and any attribute violating this condition will be discarded. The partial consolidate function will accept as parameter the summation of weights and this will denote which attributes have had a perfect match. For example, a summation value of 29 will mean facial, fingerprints and password has been successfully authenticated. The weights attached to the attributes are such that the summation combinations are unique, which makes it easy to identify the matched attributes. Similarly, a value of 17 would mean iris, facial and fingerprints alone. A similar weighting scheme for multi-modal (hybrid) biometrics has been explained in [9].

At this juncture, the consolidate function could use a one way function to generate a dynamic key just using the attributes matched. Alternatively, it could release the stored static key, which was created using all ***n*** attributes at the time of enrolment. The method can be implemented for either key generation or key release procedure. When fewer attributes are used, it might require padding of bit fields to have a secure length key. This will be an extra logic that will decide the amount of padding required based on number of attributes matched and the length of the desired key.

VII. TESTING FOR EFFICIENCY

A fine method to evaluate the performance of a biometric verification system is to present it with large number of enquiries for authorized and unauthorized users and measure the similarity rating with a stored reference. In our case, the stored reference is

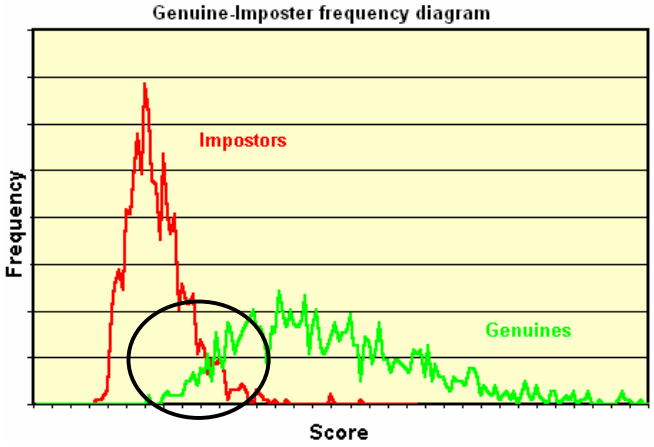

Fig 3.Genuine-Imposter Frequency diagram for single biometric authentication system

the set of digital representations for multiple biometric attributes stored during enrolment. But a match hit is made not based on one-to-one basis but on the output of consolidation. In other words, no single attribute is alone adequate to produce a match, but rather a set of attributes that produce a definitive output when presented to a hash-function.

The graph presented above, is from Biometrics FAQ [10], and shows the genuine imposter frequency diagram, mainly applicable for single-attribute biometric systems. As seen in the circled area, the two curves overlap which proves that there is a threat of an imposter gaining access. For a best case authentication, the two curves should not overlap. In single biometric authentication systems, this is not possible.

However, the system that we are investigating and endeavoring to implement will try and remove the overlapping property of the two curves. Since a perfect hit is based on multiple attributes, we believe it is possible and this will be a major basis for evaluating our system. If a means can be derived to clearly isolate the genuine users from imposters, that factor will be a key milestone that hybrid biometrics can achieve.

Another component of the test schema would be subjecting the hybrid-biometric system to attacks. Single biometric authentication systems can be exposed to attacks such as the hill-climbing method [11]. It is estimated that adding more attributes for authentication and using them in conjunction with hashing functions would reduce or possibly remove scope for attacks.

VIII. CONCLUSION

A schema for using a multitude of human physical attributes to create keys for cryptography has been presented. Combining biometrics and cryptography is a new and a significant area in

the scope of advanced communications security. The methods presented are still in an early research phase and are currently being investigated from a mathematical perspective. At the moment we are concentrating on implementation issues whilst attempting to reduce the complex overheads in managing a multitude of key generation operations using biometrics. The methods for handling hybrid biometric data seem straight forward but the intricacies of computation to embed sophisticated security is extremely compound. A major part of our investigation concerns the management of large biometric data representations and time complexity of consolidation to generate keys, hence governing the overall time requirements per authentication session.

REFERENCES

[1] Federal Financial Institutions Examination Council "Authentication in an electronic banking environment". Available www.ffiec.gov/pdf/pr080801.pdf last accessed 3rd June, 2005

[2] A. Jain, R. Bolle, and S. Pankanti, *Biometrics : Personal identification in networked society* Kulwer Academic publishers, 1999

[3] A. Burnett, A. Duffy, T. Dowling, "A biometric identity based signature scheme" Cryptology ePrint Archive, available at http://eprint.iacr.org/2004/176 last accessed 12 August, 2005

[4] G.I. Davida, Y. Frankel, B.J. Matt, "On enabling secure applications through off-line biometric identification" *Proc. IEEE Symposium Privacy and Security* pp. 148-157, 1998

[5] G.I. Davida., Y. Frankel, B.J. Matt, R. Peralta "On the relation of error correction and cryptography to an offline biometric based identification scheme" *Proc. Conf. Workshop Coding and Cryptography (WCC99)* pp. 129-138, 1999

[6] F. Hao, C.W. Chan, "Private Key Generation from On-line Handwritten Signatures," Information Management & Computer Security, Vol 10, Issue 4, pp 159-164, 2002

[7] Y. Dodis, L. Reyzin, A. Smith, "Fuzzy extractors: how to generate strong keys from biometrics and other noisy data," Eurocrypt 2004, pp 523–540, 2004

[8] A.J Menezes, P.C. van Oorschot, Vanstone S.A, *Handbook of applied cryptography* CRC press, 1997

[9] K. Shorter, I. Nice, "Biometrics and Security – An Introduction". Available http://www.qinetiq.com/home/security/securing_your_business.QuickNavPar.0007.File.pdf last accessed 9th May 2006.

[10] M. Bromba, "Bio-identification frequently asked questions". Available http://www.bromba.com/faq/biofaqe.htm last accessed 6th April 2006.

[11] A. Adler, "Vulnerabilities in biometric encryption systems" AVBPA05, 2005